BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PATENT
Docket No. 300622004600

CERTIFICATE OF MAILING BY "FIRST CLASS MAIL"

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231, on August 12, 2002.

Ruth M. Saskowski

#17
D.G.
8/28/02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the application of:

Rajesh S. GOKHALE, *et al.*

Serial No.: 09/500,747

Filing Date: 9 February 2000

For: METHODS TO MEDIATE POLYKETIDE SYNTHASE MODULE EFFECTIVENESS

Examiner: Kathleen M. Kerr

Group Art Unit: 1652

DECLARATION OF CHAITAN KHOSLA

Assistant Commissioner for Patents
Washington, D.C. 20231

Dear Sir:

I, Chaitan Khosla, declare as follows:

1. I am a Professor of Chemical Engineering and Chemistry at Stanford University, a co-founder, Board Member, and consultant of Kosan Biosciences, and a co-inventor of the above-referenced application. I have actively conducted research in the field of polyketide synthesis for over 10 years. A copy of my *curriculum vitae* is attached.

2. I understand that the specification in the above-referenced case has been criticized as failing to provide sufficient information to show that I and my co-inventors were in possession of our invention and to provide sufficient information to enable others of skill in the present art to practice the invention as claimed. The basis for this criticism is that the sequences of only a

few of the existing intramolecular and intermolecular linkers which facilitate polyketide chain transfer are disclosed, and that these sequences do not show a high degree of homology. Therefore, it is reasoned that one of ordinary skill would not be able to practice the invention either because other linkers besides those articulated are not described in the specification, or because it would not be possible for the practitioner readily to find such linkers without explicit description of them in the specification.

3. I believe this perception is in error and can verify that, as practitioners of the art, both I and my graduate students, post-doctoral students and colleagues in the field would readily be able to obtain from the published literature, familiar to us as practitioners, the desired linker structures.

4. As an example, I believe this is made clear in Chapter 2 of the thesis submission of one of my graduate students, which is appended to the response to the Office action in the above-referenced application as Exhibit B. The chapter contains a description of work demonstrating the ability of intermodular linkers to facilitate the growth of a nascent polyketide chain. This thesis was submitted November 1, 2000, and is based on publicly available information regarding polyketide sequences. All sequences from the ery, rap, FK506, rif and pik clusters were available as the date of the application to which priority is claimed, 9 February 1999; as were sequences of ole5-6 in Figure 2.10 and ole5 in Figure 2.12 . The remaining sequences were deposited in the database after that date.

5. Of particular interest are the tables set forth on pages 14-16 of this chapter, which show intrapolypeptide linkers (Figure 2.10 on page 14) and interpolypeptide linkers (Figures 2.11 and 2.12 on pages 15 and 16, respectively). As indicated in those figures, the C-termini of the ACP catalytic regions and the N-termini of the KS regions of a wide range of

already sequenced modular polyketide synthases show such a high degree of homology that identification of the linker moieties between these modules is readily apparent. We do not expect there to be substantial homology among linkers, especially with regard to the interpolypeptide linkers for which the portion bound to the C-terminus of the upstream module is shown in Figure 2.11 and the portion bound to the N-terminus of the downstream KS is shown in Figure 2.12.

6. As all of the sequences set forth in these figures are available in publications and databases of which one of ordinary skill in the relevant art would have knowledge, that person would readily be able to ascertain the amino acid sequence of either a intrapolypeptide linker or the two portions of an interpolypeptide linker based on these disclosures.

7. I understand that it is also asserted that not all possible modular polyketide synthases have yet been cloned and sequenced. Undoubtedly this is true; however, because of the striking homology exhibited by the catalytic regions in the multiple PKS whose sequences are already available, there appears no doubt that any subsequently cloned and sequenced modular PKS will exhibit a similar level of homology and be amenable to the analysis discussed above. Hence, one of ordinary skill would could identify linkers from these synthases as well.

8. For these reasons, the information requested as assertedly necessary to provide adequate disclosure of the invention as claimed in the above-referenced application is, to the extent it relates to PKS sequences known as of the priority date of the instant application and to PKS sequences later determined, readily available to, and readily accessible by, one of ordinary skill, and there would be no need to provide such skilled practitioners with any more direction than that set forth in the specification to permit successful practice of the invention as claimed with respect to such sequences.

I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further, that these statements are made with the knowledge that willful, false statements and the like so made are punishable by fine or imprisonment or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Executed this 2nd day of August, 2002, at Stanford, California

Chaitan Khosla

Chaitan Khosla

Associate Professor (b. 1964)

B.Tech., 1985, Indian Institute of Technology; Ph.D.,1990, California Institute of Technology

Camille and Henry Dreyfus New Investigator Award, 1991; National Science Foundation Young Investigator Award, 1994; David and Lucile Packard Fellowship in Science and Engineering, 1994; E. Bright Wilson Prize Lecturer, Harvard University, 1996; Linus Pauling Medal Lecturer, Stanford University, 1997; Robert Vaughan Lecturer, California Institute of Technology, 1997; Allan P. Colburn Award, American Institute of Chemical Engineers, 1997; ACS Lilly Award in Biological Chemistry, 1999; NSF Alan T. Waterman Award, 1999 Organic and Biophysical Chemistry

650-723-6538
ck@chemeng.stanford.edu

Principal Research Interest

Biocatalysis

Our research interests, which straddle the interface of chemistry, biology, and engineering, revolve around microbial natural products. We work on three themes pertaining to natural products:

(1) *Biosynthetic mechanisms.* We are interested in understanding fundamental mechanisms underlying the biosynthesis of natural products. Of particular interest are multi-enzyme systems (a.k.a. megasynthases) which catalyze the biosynthesis of complex natural products such as polyketides and nonribosomal peptides. These multifunctional catalysts are remarkable in their ability to chaperone highly reactive biosynthetic intermediates through 10-100 distinct catalytic events. Individually they are exquisitely selective with respect to product functionality and stereochemistry; yet as a family of enzymes, they generate breathtaking diversity. Examples of our laboratory*s contributions to the current understanding of polyketide biosynthetic mechanisms can be found in some of the references below.

(2) *Engineered biosynthesis of "unnatural" natural products.* Over the past three decades breakthroughs in molecular genetics have propelled a revolution in our ability to analyze and manipulate the structures and properties of nucleic acids and proteins. Our research is premised on the notion that nature has evolved similar genetic codes for the biosynthesis of polyketides and nonribosomal peptides. However, in contrast to nucleic acids and proteins, whose structures are directly encoded within modular templates, the biosynthesis of these molecules lacks the involvement of overt template molecules. Instead, it appears that molecular diversity arises from the modularity of these megasynthases at the domain/subunit level. We seek to develop novel, interdisciplinary methods for exploiting this modularity for the practical synthesis of new natural product analogs. Again, examples of our accomplishments can be found in some of the references below.

(3) *Biology of natural and "unnatural" natural products.* The ability to practically manipulate the structures of natural products offers the opportunity to understand their multi-faceted mechanisms of action. Of particular interest at present to our laboratory is the mechanism of action of apoptolidin, a newly discovered natural product that acts as a potent and selective inducer of programmed cell death (apoptosis).

Representative Publications

1 "The Chemistry and Biology of Fatty Acid, Polyketide, and Non-Ribosomal Peptide Biosynthesis," C.W. Carreras, R. Pieper, and C. Khosla, *Topics Curr. Chem.*, 188, 85-126 (1997).

2 "Harnessing the Biosynthetic Potential of Modular Polyketide Synthases," C. Khosla, *Chem. Rev.*, 97, 2577-2590 (1997).

3 "Harnessing the Biosynthetic Code. Combinations, Permutations, Mutations," D.E. Cane, C.T. Walsh, and C. Khosla, *Science*, 282, 63-68 (1998).

4 "Tolerance and Specificity of Polyketide Synthases," C. Khosla, R.S. Gokhale, J.R. Jacobsen, and D.E. Cane, *Ann. Rev. Biochem*, in press (1999).

Last modified: Fri Nov 19 02:18:12 PST 1999

Chapter 2

QUALITATIVE ANALYSIS OF HETEROLOGOUS MODULE FUSIONS

Exhibit B

2.1. INTRODUCTION

In the absence of structural information or a model of intermodular interfaces, investigation of chain transfer between modules presented a challenge. The simplest context in which transfer occurs is the truncated DEBS1-TE construct (Figure 2.1). However, instead of dissecting a natural bimodular protein, such as DEBS1, a panel of heterologous module-module fusions were constructed. By observing the successful fusions (as judged by polyketide production), it was anticipated that important factors for chain transfer could be discovered and analyzed.

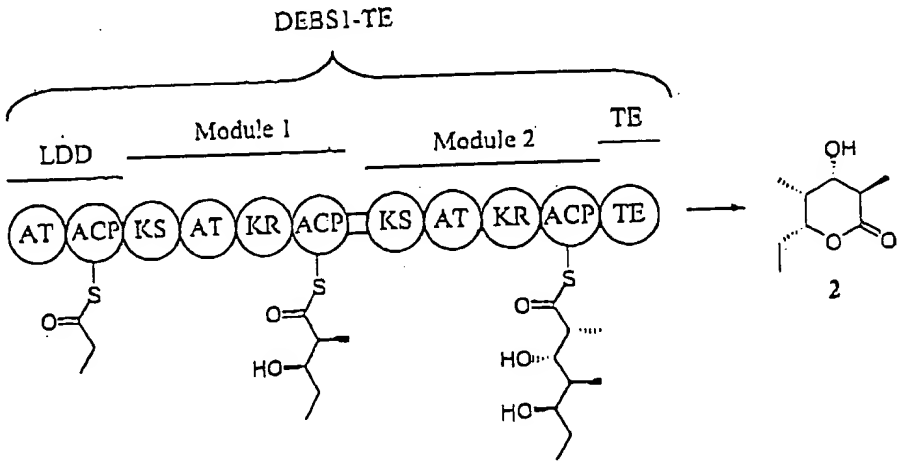

Figure 2.1 DEBS1-TE. Truncated form of DEBS, DEBS1-TE requires intermodular transfer between modules 1 and 2 to yield the triketide lactone product.

Although active site domains within modules[23,24] and the loading didomain of DEBS[22] have been successfully replaced through genetic manipulation and despite ample evidence of the functional independence of modules[17,19], successful module fusions had never been reported. Therefore, judicious planning was needed to select modules and fusion sites best suited for heterologous transfer. Because it was assumed that selectivity would be dictated by the recipient module, the first step required finding modules that demonstrated tolerance towards unnatural substrates. At the time of these experiments, preliminary work had shown that individually purified modules 2, 5, and 6 of DEBS, each fused with the terminal thioesterase (M2-TE, M5-TE, M6-TE), were capable of

recognizing and elongating the *N*-acetylcysteamine (NAC)-diketide thioester **3** (NDK) to produce the triketide lactone **2** (Figure 2.2A)[27]. Similarly, module 3 fused to the thioesterase (M3-TE) extended NDK to the unreduced triketide ketolactone **4** (Figure 2.2B). Since NDK mimics the natural substrate of only M2-TE, modules 3, 5, and 6 each exhibited relaxed substrate specificity by accepting NDK. These *in vitro* experiments immediately intimated a framework for studying module fusions: replacement of M2 in DEBS1-TE with M3, M5, and M6 to generate LDD-M1-M3-TE, LDD-M1-M5-TE, and LDD-M1-M6-TE (where "LDD" represents the loading didomain), respectively. The concern regarding substrate specificity was tempered because the modules would be presented with the same NDK substrate, except it would be via module 1, rather than via exogenous loading.

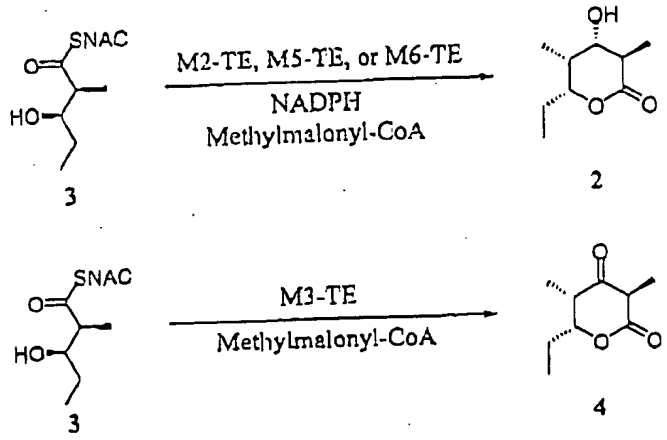

Figure 2.2 Individual module assays. (A) Exogenously fed NDK is accepted and extended by modules 2, 5, and 6 to form the triketide lactone. (B) Module 3 fused to the thioesterase also accepts NDK, but instead catalyzes production of the unreduced triketide ketolactone.

These bimodular constructs presented attractive targets, but previous attempts at making these fusions yielded proteins unable to catalyze *in vivo* the biosynthesis of the expected triketide lactones **2** (in the case of LDD-M1-M6-TE)[28] or **4** (in the case of LDD-M1-M3-TE)[29]. However, the purified proteins were able to extend NDK to their respective triketide lactones, thereby demonstrating the catalytic competence of the downstream fused modules[30] (Figure 2.3).

Based upon these results, we hypothesized that in each construct, both module 1 and the transplanted module retained their individual activities, but simply lacked the proper connectivity required to allow transfer between them.

By selecting a different site for fusion, the LDD-M1-M3-TE and LDD-M1-M6-TE constructs were recloned in an active form, reflected by their ability to efficiently catalyze the biosynthesis of the expected triketides *in vivo* in *Streptomyces coelicolor*. This surprising result implicated the importance of short "linker" regions located outside the conserved catalytic domains. Other heterologous fusions exploiting this understanding suggested the true "modularity" of modules, where modules may be replaced within a framework of linkers that maintain the overall organization.

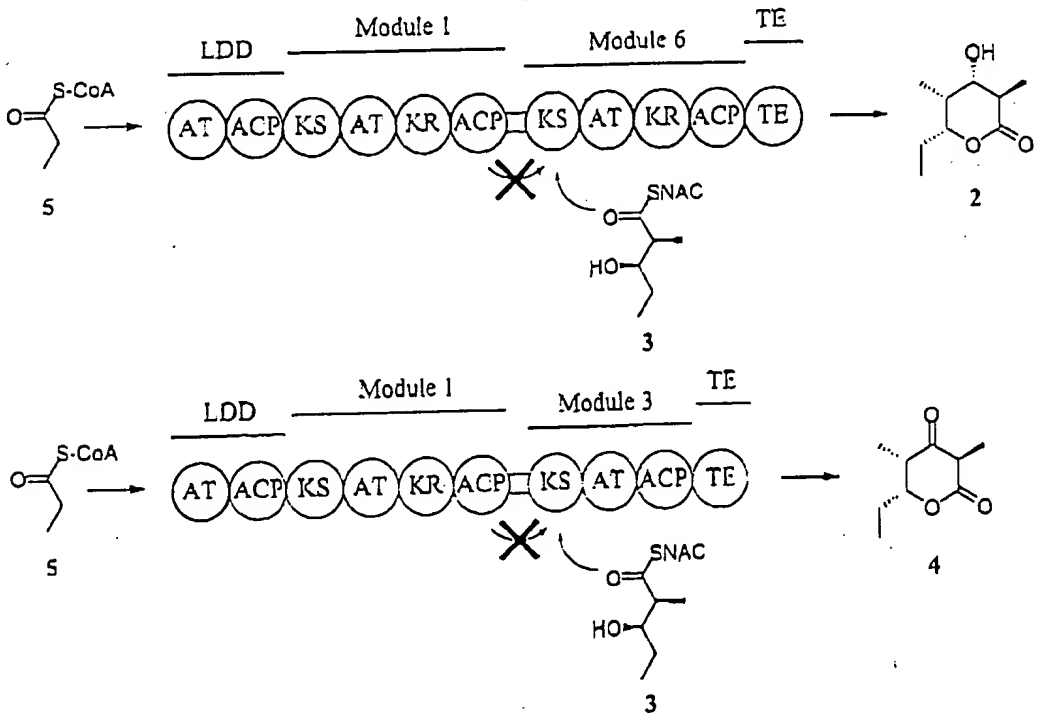

Figure 2.3 Previous attempts of bimodular fusion. Although the bimodular constructs were not able to produce triketides from their natural substrate, propionyl-CoA 5, the downstream module was able to process exogenous NDK to the expected triketide lactones.

2.2. Materials and Methods

2.2.1. Construction of Module Fusions for *S. coelicolor* Expression

I. LDD-M1-M3-TE (pST97): This bimodular fusion contains module 1 fused to

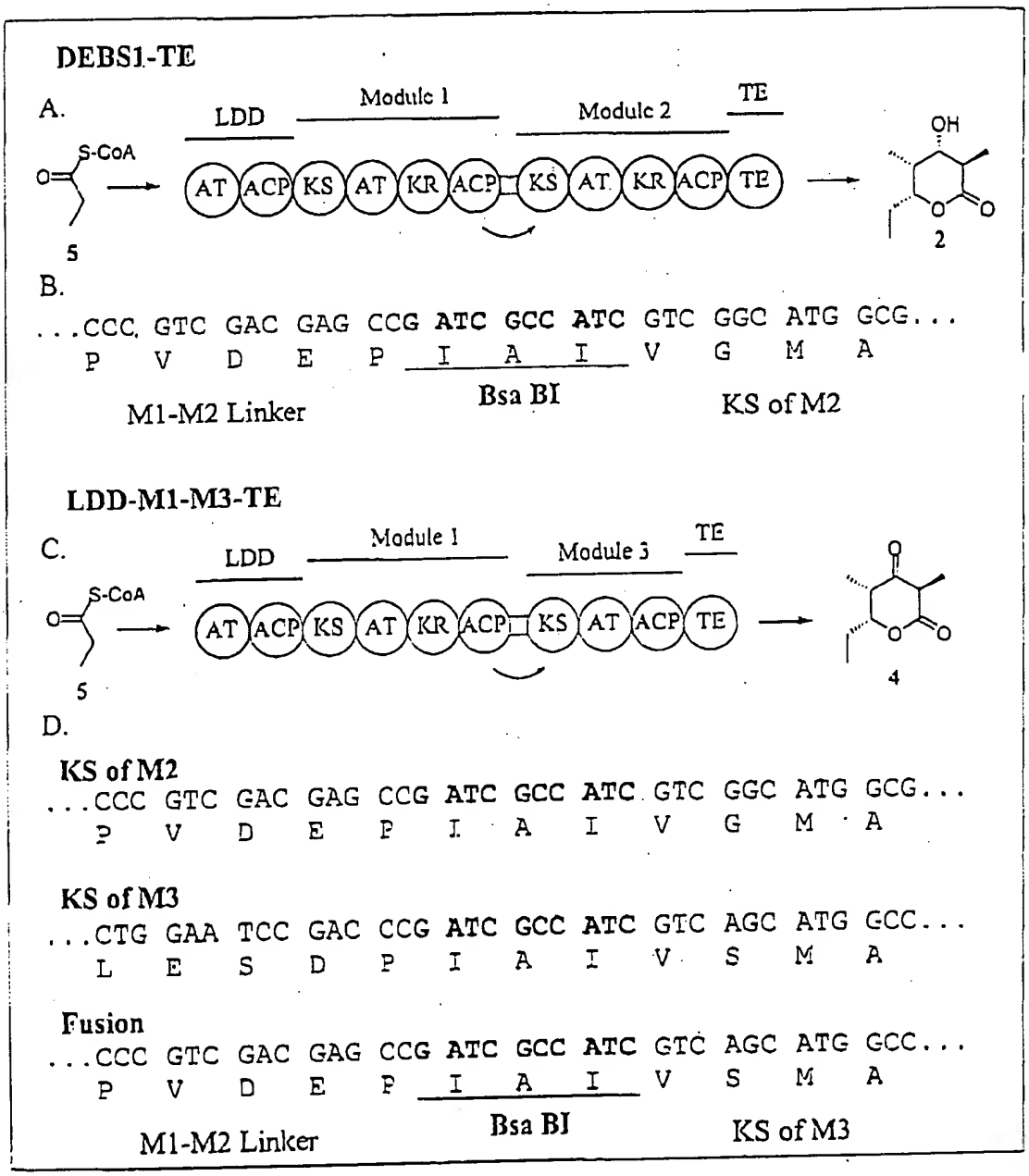

Figure 2.4 DEBS1-TE and LDD-M1-M3-TE. (A) Schematic diagram showing the major polyketide product. Curved arrows represent where intermodular transfer occurs. (B) Nucleotide and amino acid sequence at the intermodular junction. (C) Schematic for LDD-M1-M3-TE. (D) Nucleotide and amino acid sequences near the splicing site of DEBS M2 KS and M3 KS. The composite sequences are shown below the contributing sequences. Nucleotides and amino acids of the enzyme recognition site are highlighted in bold font.

module 3 by taking advantage of the natural, conserved Bsa BI site located at the start of the KS domain. Thus, the "linker" sequence bridging the fused modules is the natural sequence between modules 1 and 2, as in DEBS1. Module 3 fused to the thioesterase

domain was supplied as a Bsa BI-Eco RI fragment from a derivative of pRSG34[27]. This fragment was cloned behind M1 in a derivative of pCK12[20] to generate pST97[27] (Figure 2.4 C and D).

II. LDD-M1-M5-TE (pST142): This bimodular fusion was constructed in an identical manner as pST97, except it contains module 5 instead of module 3. Thus, it also maintains the natural M1-M2 linker between modules 1 and 5. Module 5 fused to the thioesterase domain was supplied as a Bsa BI-Eco RI fragment from a derivative of pRSG46[27]. This fragment was cloned behind M1 in a derivative of pCK12 to generate pST142 (Figure 2.5 A and B).

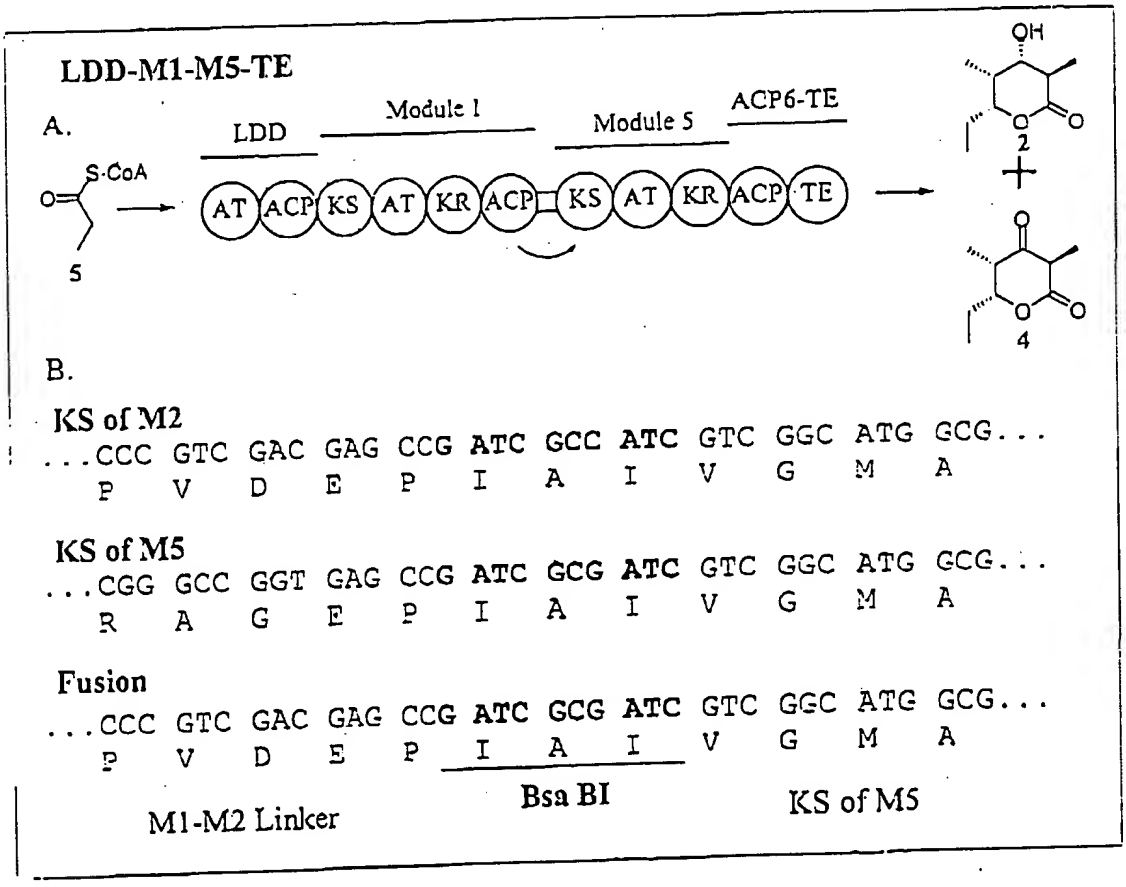

Figure 2.5 LDD-M1-M5-TE. (A) Schematic diagram showing the major polyketide product. Curved arrow represents where intermodular transfer occurs. (B) Nucleotide and amino acid sequences near the splicing site of DEBS M2 KS and M5 KS. The composite sequences are shown below the contributing sequences. Nucleotides and amino acids of the enzyme recognition site are highlighted in bold font.

III. LDD-M1-M6-TE (pST96): This bimodular fusion was constructed with the same cloning strategy described for pST97 and pST142. Module 6 with the thioesterase domain was supplied as a Bsa BI-Eco RI fragment from a derivative of pJRJ10[25]. This fragment was cloned behind M1 in a derivative of pCK12 to generate pST96 (Figure 2.6 A and B) [27].

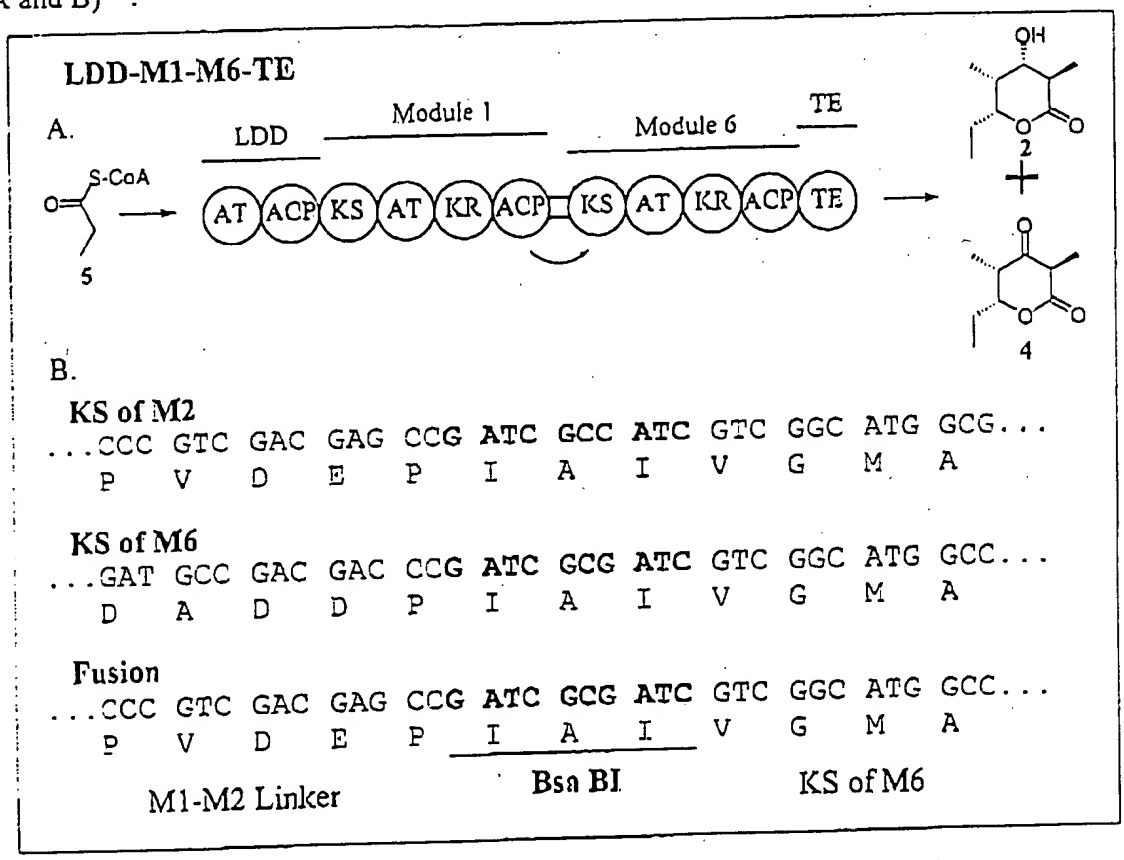

Figure 2.6 LDD-M1-M6-TE. (A) Schematic diagram showing the major polyketide product. Curved arrow represents where intermodular transfer occurs. (B) Nucleotide and amino acid sequences near the splicing site of DEBS M2 KS and M6 KS. The composite sequences are shown below the contributing sequences. Nucleotides and amino acids of the enzyme recognition site are highlighted in bold font.

IV. LDD-M1-rifM5-TE (pST114): Module 5 of the rifamycin polyketide synthase[31,32] (rifM5) contained the conserved Bsa BI site upstream of its β-ketoacyl-acyl carrier protein synthase (KS) domain, so the site of module-module fusion required no engineering. However, to attach the thioesterase domain to the C-terminus of rifM5, an

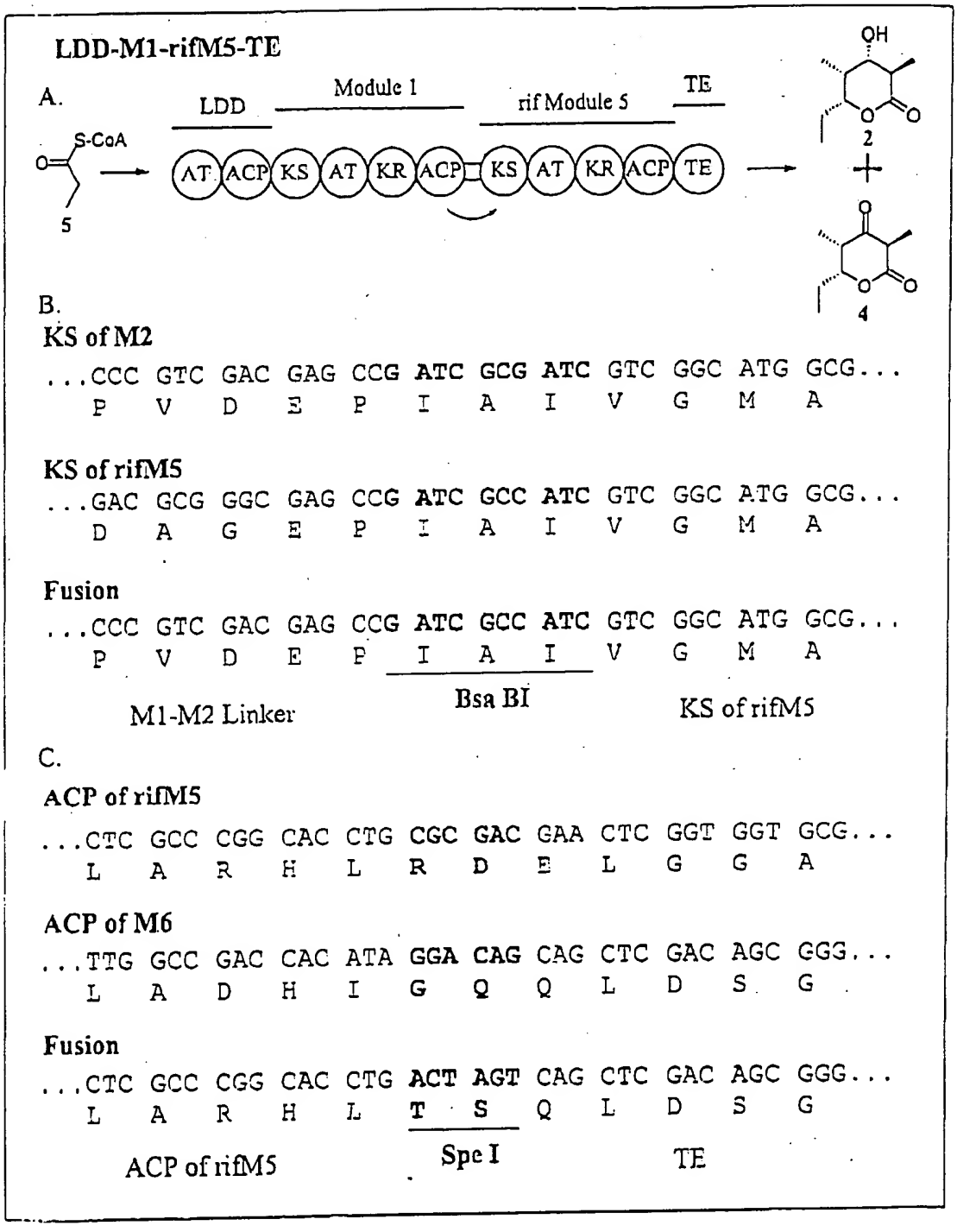

Figure 2.7 LDD-M1-rifM5-TE. (A) Schematic diagram showing the major polyketide product. Curved arrow represents where intermodular transfer occurs. (B) and (C) Nucleotide and amino acid sequences near the splicing site of rifM5 KS and DEBS M2 KS, and rifM5 ACP and DEBS M6 ACP, respectively. The composite sequences are shown below the contributing sequences. Nucleotides and amino acids that were changed for enzyme recognition site engineering are highlighted in bold font.

Spe I site was required. The Spe I site was placed downstream from its acyl carrier protein (ACP) domain, by using the primers 5'-**CAA CCG CAT GCG GAC CGC**-3' and 5'-TTC ACT AGT **CAG GTG CCG GGC GAG**-3' (nucleotides in boldface are complementary to the rifamycin sequence) to amplify the region between a natural Sph I site and the engineered Spe I site. The location of the Spe I site was chosen by aligning the amino acid sequences of the ACP domains of the six modules of DEBS and rifM5; the Spe I was placed in the region homologous with the engineered DEBS1-TE (pCK12) [20] and DEBS1-M3-TE (pCK13) [18]. A three fragment ligation was performed with the above PCR amplicon, a Bsa BI-Sph I fragment derived from pHu72 that contained the remaining upstream portion of rifM5, and a plasmid containing the thioesterase domain to generate pST100. A Bsa BI-Eco RI fragment from pST100 containing rifM5-TE was inserted behind M1 in a derivative of pCK12 to yield pST114 (Figure 2.7 A, B, and C) [27].

V. M1-rifM5-DEBS2-DEBS3 (pST113): The DEBS1 portion of a derivative of pCK7[26] was replaced with the homologous region from pST100 using the Pac I and Spe I restriction sites to make pST113 (Figure 2.8 A and B) [27]. The resulting plasmid contains rifM5 substituted for module 2 of DEBS. The linker between module 1 and rifM5 is identical to those used in the previous constructs, but in addition, this construct also maintains the natural linker region between modules 2 and 3. Therefore, in place of the thioesterase domain attached to its C-terminus, rifM5 has the C-terminus of module 2 of DEBS.

2.2.2. Isolation of Polyketides Produced *In vivo*

To transform *S. coelicolor* with the above plasmids, the requisite non-methylated DNA was obtained via transformation into *E. coli* ET12567[33]. The resultant DNA was transformed into *S. coelicolor* CH999 following established protocols[34]. Transformants were grown on R5 medium at 30° and selected with an overlay of thiostrepton (50 g/L in DMSO) in 1 mL of water (500 mg/L final concentration). Multiple transformants of each construct

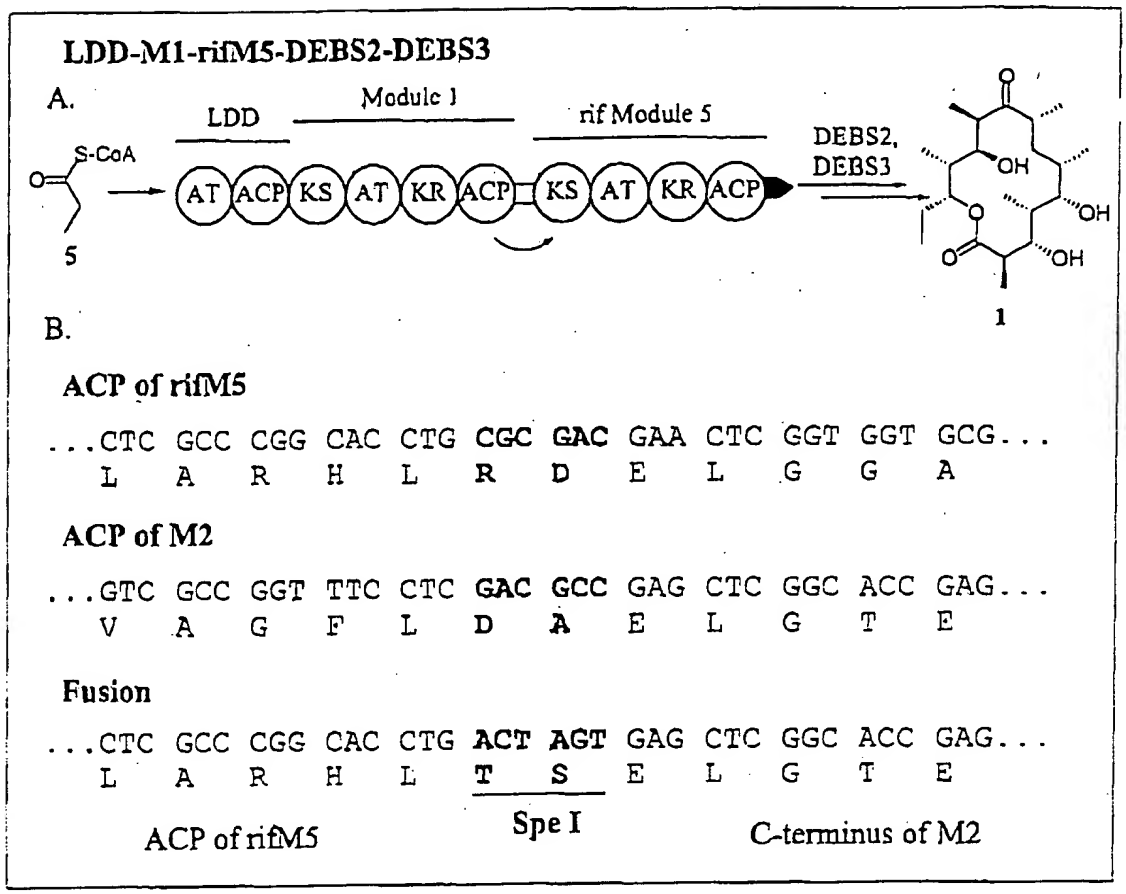

Figure 2.8 LDD-M1-rifM5-DEBS2-DEBS3. (A) Schematic diagram showing the major polyketide product. Curved arrow represents where intermodular transfer occurs in first polypeptide before transfer to DEBS2 and DEBS3. (B) Nucleotide and amino acid sequences near the splicing site of rifM5 ACP and DEBS M2 ACP. The composite sequences are shown below the contributing sequences. Nucleotides and amino acids that were changed for enzyme recognition site engineering are highlighted in bold font.

were re-streaked onto R5 plates supplemented with sodium propionate (500 mg/L) and thiostrepton (50 mg/L) and were grown into confluent lawns (7 days). The agar medium was then homogenized and extracted three times with warm (40°C) ethyl acetate. The solvent extract was dried over magnesium sulfate, concentrated, and filtered through silica gel (1.2 x 4 cm, eluted isocratically with 50% ethyl acetate in hexanes). Evaluation of the crude extracts was performed via thin layer chromatography (TLC) in 50% ethyl acetate in hexanes and stained with vanillin. Purification of the products was carried out as described below.

2.3. RESULTS

2.3.1. Production of triketide ketolactone by CH999/pST97

S. coelicolor CH999/pST97 produced triketide ketolactone 4 (ca. 10 mg/L), as determined by ^{1}H NMR spectroscopy (Figure 2.4 C). As mentioned earlier, this product was the same product detected when NDK, the NAC-thioester analog of the diketide intermediate produced by module 1, was incubated with purified M3-TE[27]. The ketolactone was purified via silica gel chromatography (1.2 x 10 cm, eluted isocratically with 40% ethyl acetate in hexanes) and had an R_f of 0.5 in 50% ethyl acetate in hexanes.

2.3.2. Production of triketide lactone by CH999/pST142

S. coelicolor CH999/pST142 produced triketide lactone 2 (ca. 5 mg/L) and ketolactone 4 (ca. 2 mg/L) as determined by ^{1}H NMR spectroscopy (Figure 2.5 A). The reduced lactone and the ketolactone were identical to those produced in the assays of NDK with M5-TE and M3-TE, respectively. The products were corroborated by unit mass spectrometry (ESI). M+ calculated for $C_9H_{16}O_3$ and $C_9H_{14}O_3$: 172 and 170. Observed: 172 and 170. MNa+ calculated for $(C_9H_{16}O_3)Na$ and $(C_9H_{14}O_3)Na$: 195 and 193. Observed: 195 and 193. The triketides were purified via silica gel chromatography (1.2 x 10 cm, eluted isocratically with 50% ethyl acetate in hexanes). Inspection using ^{1}H NMR indicated that the collected fractions were not pure. Further purification was obtained by running a preparative TLC plate in 50% ethyl acetate in hexanes and using the 254 nm absorbance of the ketolactone as a marker. R_f estimates of the triketide lactone and ketolactone in the solvent system were 0.45 and 0.5, respectively.

2.3.3. Production of triketide lactone by CH999/pST96, CH999/pST114

S. coelicolor CH999/pST96 and CH999/pST114 produced triketide lactone 2 and triketide ketolactone 4 (ca. 5 mg/L), as determined by ^{1}H NMR spectroscopy (Figure 2.6,

Figure 2.7). Purification was carried out as for CH999/pST97. The products were identical to those extracted from CH999/pST142.

2.3.4. Production of 6-deoxyerythronolide B by CH999/pST113

S. coelicolor CH999/pST113 produced 6-deoxyerythronolide B 1 (6-dEB) (ca. 2 mg/L), as determined by ^{1}H NMR spectroscopy (Figure 2.8 A). The product was also confirmed by mass spectrometry (ESI). MH+ calculated for $(C_{21}H_{33}O_6)H+$: 387. Observed: 387. 6-dEB was purified using silica gel chromatography (1.2 x 10 cm, gradient of 25-50% ethyl acetate in hexanes). Observed $R_f = 0.5$ in 50% ethyl acetate in hexanes.

2.4. Discussion

The ability of the fused modules to catalyze the biosynthesis of the expected polyketide products cogently demonstrates the feasibility of manipulating the organization of modules in PKSs. The robust productivity of these engineered PKSs, relative to domain substituted PKSs[15,35], also reflects the advantage of keeping modules as intact catalytic units. More importantly, however, these results highlight the crucial role of the linker regions located between modules. Prior to this discovery, the significance of the non-conserved regions containing the linkers had not been recognized, exemplified by the introduction of mutations in earlier attempts of the M1-M3-TE and M1-M6-TE module fusions (Figure 2.9). Inspection of the aligned sequences indicates that the changes caused

```
eryM1-M2       LAAHLAAELG GATGAEQAAP A...TT.APV DEPIAIVGM
eryM1-*M3      LAAHLAAELG GATGAEQAAP A...TT.APA SEPIAIVSM
eryM1-*M6      LAAHLAAELG GATGAEQAAP A...TT.APA SEPIAIVGM
eryM1-M3       LAAHLAAELG GATGAEQAAP A...TT.APV DEPIAIVSM
eryM1-M5       LAAHLAAELG GATGAEQAAP A...TT.APV DEPIAIVGM
eryM1-M6       LAAHLAAELG GATGAEQAAP A...TT.APV DEPIAIVGM
eryM1-rifM5    LAAHLAAELG GATGAEQAAP A...TT.APV DEPIAIVGM
```

Figure 2.9 Alignment of heterologous intermodular fusions. Alignment of amino acids showing the mutations made in the inactive bimodular constructs (indicated with *) and the subsequent restoration in the active constructs.

by the introduction of a cloning site upstream from the M3 and M6 KS domains deleteriously affected the channeling of intermediates between the modules. Reversing the VD to AS mutation and use of the conserved Bsa BI site established the connectivity needed for transfer, as reflected by the efficient production of the expected triketide products. Furthermore, the successful transplantation of rifM5 into the bimodular framework suggests that this fusion strategy may apply generally, especially given that the Bsa BI site is conserved in 4 of 6 and 8 of 10 modules in DEBS and rifamycin, respectively.

The impact of the VD to AS alteration was greater than anticipated, especially considering that M3 and M4 of DEBS both contain a serine at the second position and M5 and M6 of DEBS both contain alanine at the first position. However, the identity of either amino acid does not appear to be as important as the presence of an acidic side chain. Analysis of the homologous amino acids for modules similarly attached to an upstream module, including M2, M4, and M6 of DEBS and modules from the rifamycin, epothilone, avermectin, oleandomycin, FK506, rapamycin, and pikromycin polyketide synthases, indicates that in nearly every case, one of the two amino acids is aspartate or glutamate (Figure 2.10). Although this trend certainly does not prove the necessity of the acidic group, it does suggest, in concert with the experimental results, that the composition of the intermodular linker region is not arbitrary and does contain important features for chain transfer.

Whereas the bimodular constructs required transfer across only one heterologous junction, the production of 6-dEB by pST113 reflected the ability of rifM5 to accept and donate intermediate chains across two heterologous junctions. Transfer across the second junction, between rifM5 and M3 of DEBS, represented a fundamentally different phenomenon from that observed in the bimodular constructs. The transfer in the bimodule fusions was expected to tolerate some degree of protein change since the covalent attachment ensured the close proximity of the modules, even if specific protein-protein interactions were mitigated. However, the interaction between rifM5 and M3 did

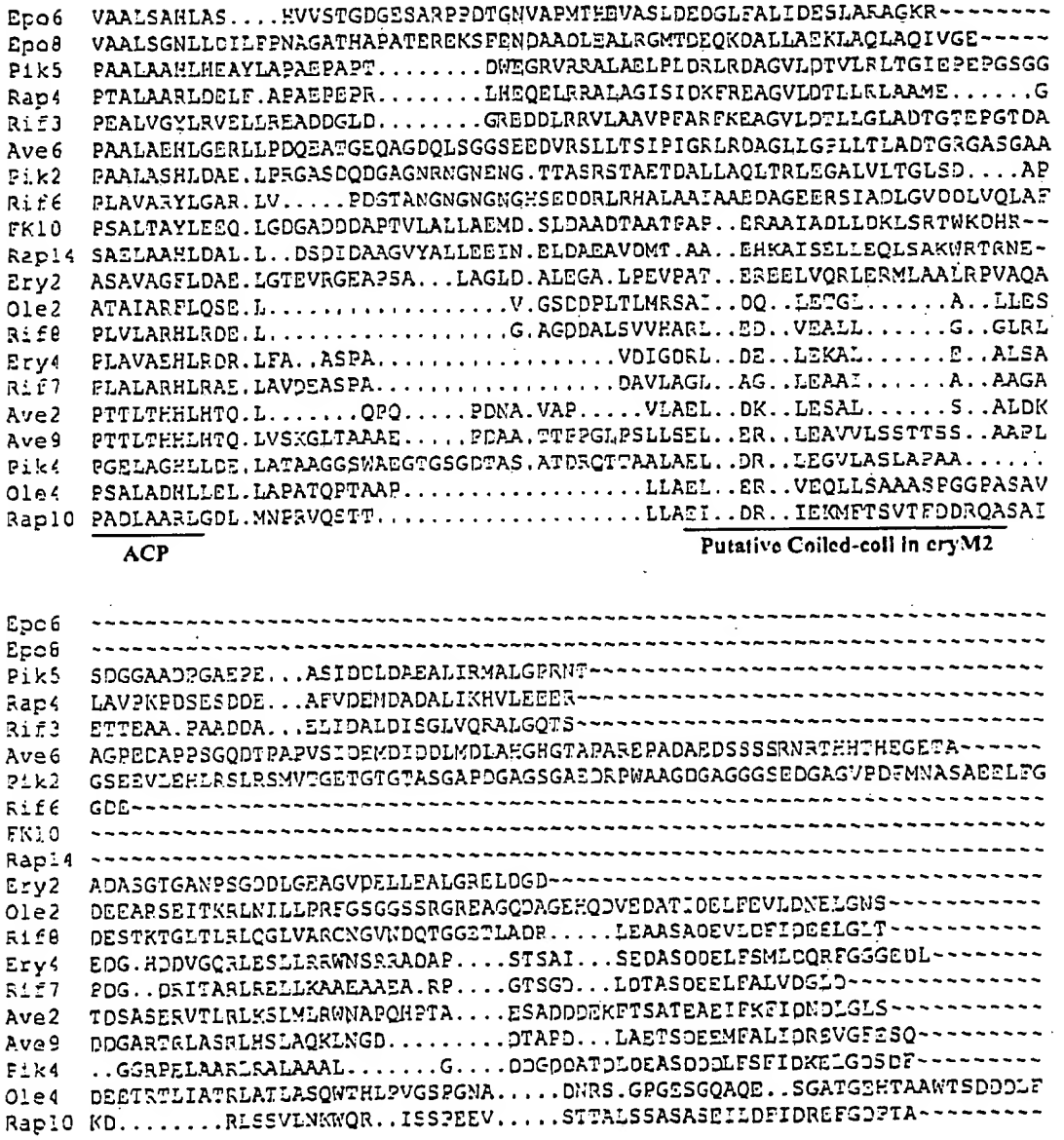

Figure 2.11 Alignment of C-terminal interpolypeptide linkers.
Alignment of regions following the ACP domain (indicated by label) of C-terminal modules in DEBS (Ery), rifamycin (Rif), epothilone (Epo), avermectin (Ave), oleandomycin (Ole), FK506 (FK), rapamycin (Rap), and pikromycin (Pik). Sequences are grouped according to the degree of homology, according to PileUp (Genetics Computer Group). The region predicted to contain a coiled-coil motif is also underscored (corresponding to the sequence of EryM2). The positively and negatively charged residues are in blue and red, respectively.

appear to play a role in the protein-protein recognition for chain transfer between modules on separate polypeptides. Figure 2.11 shows the C-terminal linker region of M2 along with the homologous region of modules from the above-mentioned PKSs. Whereas the intrapolypeptide linkers demonstrated obvious homology, especially within

a given PKS (e.g. rapamycin), the C-terminal interpolypeptide linkers did not seem retain much homology. Comparison of the ordering of the sequences suggested that even linkers from the same PKS had few features in common. The diversity of these linkers may reflect their suspected ability to selectively interact with other polypeptides. Interestingly, their putative partners, the N-terminal interpolypeptide linkers, contain relatively high homology in comparison. Across the different PKSs, the pattern of

```
Epo7    ----------MIDREGQLLERLREVTLALRKTLNERDTLE...LEKTEPIAI
Epo9    ---------MATTNAG....KLEHALLLMDKLAKKNASLE...QERTEPIAI
Rap5    ------------MREDQLLDALRKSVKENARLRKANTSLRAAMD...EPLAI
Rif4    MANQSWRKNMSAPNEQIVDALRASLKENVRLQQENSALAAAAA...EPVAI
Pik6    ----------MTSSNEQLVDALRASLKENEELRKESRRRADRRQ...EPMAI
Ave7    ----------MDTSSEKLVDALRASLKANQTLRARNEQLAAAMEASSEPIAI
Rif7    ---------MVSASYEKVVEALRKSLEEVGTLKKRNRQLA...DAAGEPIAI
Ery3    ----------MTDSEKVAEYLRRATLDLRAARQRIRELES......DPIAI
Ole3    ----------MTNDEKIVEYLKRATVDLRKARHRIWELED......EPIAI
Ery5    ----MSGDNGMTE.EKLRRYLKRIVTELDSVTARLREVEHR...AGEPIAI
Ole5    ----------MAEAEKLREYLWRATTELKEVSDRLRETEER...AREPIAI
Ave10   ----------MA.NEEKLRDYLKRVTADLLNVRRRLQQIESG...EQEPIAI
Ave3    -------MQLA.NEAKLLEYLKRVTADLDRTRRRLYEVVER...EQEPIAI
Pik5    ----------MANNEDKLRDYLKRVTAELQQNTRRLREIEGR...THEPVAI
Pik3    ---------MSTVNEEKYLDYLRRATADLEEARGRLRELEAK...AGEPVAI
Rif8    ----------MADEGQLRDYLKRAIADARDARTRLREVEEQ...AREPIAI
FK7     ----------MPEQDKTVEYLRWATTELQRTRAEL....AA..HS.EPLAI
Rap11   ----------MPEQDKVVEYLRWATAELHTTRAKLEALAAA..NT.EPLAI
Rif9    ----------MATDEKLLKYLKRVTAELHSLRKQ....GAR..HADEPLAV
Epo3    ---------MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAI
                                                  Putative coiled-coil   KS
```

Figure 2.12 Alignment of N-terminal interpolypeptide linkers. Alignment of regions preceding the KS domain (indicated by label) of N-terminal modules in DEBS (Ery), rifamycin (Rif), epothilone (Epo), avermectin (Ave), oleandomycin (Ole), FK506 (FK), rapamycin (Rap), and pikromycin (Pik). Sequences are grouped according to the degree of homology, according to PileUp (Genetics Computer Group). The region predicted to contain a coiled-coil motif is also underscored (corresponding to the sequence of EryM3). The positively and negatively charged residues are in blue and red, respectively.

charged residues remained consistent. However, like the C-terminal linkers, the modules within a given PKS do not seem to cluster as closely as for the intrapolypeptide linkers. The high proportion of charged residues on the interpolypeptide linkers implicate their exposure to solvent and may also serve as a means of specific recognition. This feature is analyzed in greater detail in Chapters 4 and 5.

The results of these *in vivo* experiments revealed the heretofore unrealized importance of the short, non-conserved linker regions. Although proper linker maintenance appears to facilitate functional reorganization of modules, the catalytic

```
Epo4-5      FDHPTVQRLVEHLLVDVL...KLEDR..SDI.............QHVRSLASQEP.IAI
Epo7-8      FDHPTVERLVAHLLTQVL...KLEDR..SDT.............RHIRSVAADDD.IAI
Epo3-4      FDHPTVERLVEYLLSQAL...ELQDR..TDV.............RSVRLPATEDP.IAI
Epo5-6      FDHPTVDALTRWLLDKVL...AVAEPSVSPA.............KSSPQVALDEP.IAV
Ave5-6      FDYPTPQECAAHLRTQLV...DLDDEEDAALS.......NALPQVAHRRTVEDEP.IAI
Ole1-2      FDYARPAALAGHLRSRLI.....DDDGD...H.......GALPGV..EKHAIDEP.IAI
FK9-10      FSHPTAEALATHLLELL.....D.APAALTTGAPLPAVTAAPGTAARDQ...DEP.IAI
Rap13-14    FDYPTANALAAHLLGKL.....DIPPVQQRLEAPAPSTVTGPADPVADPPSANEP.IAI
Ery3-4      FDHPTVTALAQHLRARLV...G..EADQ..AAVRV...VGAA........DESEP.IAI
Ery5-6      FDHPTITRLADHYLERLV...GAAEAEQAPALVRE...VPPK........DADDP.IAI
Ave3-4      FDYPTPMALCQFLRAAI.........VGADTGTTTRLPLTAV........PADEP.IAI
Ave4-5      FDYPSPTKLAQFLLSEI.........AEPQPDNSTPLPRPRA.....,...ELDEP.IAI
Rap11-12    FDYPTPAALAARL.DELF...TGENPV......PVRGPVSAV........AQDEP.LAI
Rap9-10     FDYPTPTALAARL.DELF...TGENPA......PVRGPVSAV........AQDEP.LAI
Rap2-3      FDYPTPTALAVRL.GELF...TGENPV......PVRGPVSAV........AQDEP.LAI
Rap6-7      FDYPTPTALAARL.DELF...TGENPA......PVREPVPAV........AQDEP.LAI
Rap1-2      FDYPTPAALAARL.EELF...TGENPA......PVRTSVSVV........AQDEP.LAI
Rap3-4      FDYPTPTVLAARL.DELF...TGENPA......PVRGPVSVV........GQDEP.LAI
Rap8-9      FDYPTPGALAARL.EELF...TGENPV......QVRTPVSAV........GQDEP.LAI
Rap7-8      FDYPTPTALAARL.GEWF...VGETPV......PVRTSVSVV........AQDEP.LAI
Rap12-13    FDYPTPAILATRL.GELF...TGENPA......PVRPSVSVV........GQDEP.LAV
Rap5-6      FDYPTPAVLAARL.GELF...TGENPV......LVR.TASVV........GQDEP.LAI
FK7-8       FDPPTPRALAARLGDELA...GTRAPV......AARTAATAA........AHDEP.LAI
FK8-9       FDHPTPRVLAAELRTDLF...GTAAPL......PARTART..........RHDEP.LAI
Rif4-5      FDYPTPVALARHLREEL.....GETVAGAPATPV....TTVA........DAGEP.IAI
Rif9-10     FDYPTPVALARYLRDEL.....GDTVATTPVA.T....AAAA........DAGEP.IAI
Rif5-6      FDYPNPQALARHLRDEL.....GG.AAQTPVTTA....AAKA........DLDEP.IAI
Ave10-11    FDYPTPSALAGYLKEQLE...EGA...GGQRDIAPPVPASRV........DVDEP.IAI
Ave11-12    FDHPRPVALAAHIHEQLS...GGSPTTGTALALALRAPAPRV........DVDEP.IAI
Ave7-8      FDHPNPTTLTHHLHTQLL...GSDSTASIP...APRAAAVPA........DQDEP.VAI
Ave8-9      FDQPNAATLARHLRRELM...GDDAEGETP...S.QVALHQV........AADEP.IAI
Ery1-2      FDHPDVRTLAAHLAAELG...G..ATGAEQ...AAPATTASV..........DEP.IAI
Ave1-2      FDHPTPAKLAVHLQNQLR...G....TAAESAPSAAAVTATA.........SVTEP.IAI
Pik3-4      FDYPTPRTLAEFLLAEIL...GEQAGAGEQLPV.DGGV.............DDEP.VAI
Rif2-3      FDYPNPAVLADHLRAELL...GERPAA..PAPV.TRDV..............SDEP.IAI
Rif1-2      FDYPKPPALADHLRAKLF...G..SAANRPAEI.GTAA..............AEEP.IAI
Ole3-4      FDYPTTNGLAEYLRSELF...GVSGAPADLSVVRNADE..............EDDP.VVI
Pik1-2      FDHPTPLALVSLLRSEFL...GDEETADARRSAALPATVGAGAGAGAGTDADDDP.IAI
Ole5-6      FDHPNATAIARFLQSQLLPDAESESAVPSSPEDEVRQALASLSLDQLKGAGLLDPLLAL
                ACP                                                  KS
```

Figure 2.10 Alignment of intrapolypeptide linkers. Alignment of regions linking covalently attached modules in DEBS (Ery), rifamycin (Rif), epothilone (Epo), avermectin (Ave), oleandomycin (Ole), FK506 (FK), rapamycin (Rap), and pikromycin (Pik). Sequences are grouped according to the degree of homology, according to PileUp (Genetics Computer Group). The intrapolypeptide linker sequences are flanked by the consensus ACP and KS sequences, as indicated by the labels under the list. The positively and negatively charged residues are in blue and red, respectively.

not contain such a crutch since they reside on different polypeptides. Therefore, the observed production of 6-dEB implied that some recognition of rifM5 and M3 was retained. Because the C-terminus, fused to rifM5, was the only feature remaining from M2, the ca. 90 amino acid linker region was implicated in the recognition. Thus, not only do the linkers appear essential for chain transfer between fused modules, but they also